3-2012

Post-Acute Care Payment Reform Demonstration: Final Report Volume 1 of 4

Barbara Gage

Melissa Morley

Laura Smith

Melvin Ingber

Anne Deutsch

See next page for additional authors

Authors
Barbara Gage, Melissa Morley, Laura Smith, Melvin Ingber, Anne Deutsch, Tracy Kline, Jill Dever, Judith Abbate, Richard Miller, Brieanne Lyda-McDonald, Cynthia Kelleher, Danielle Garfinkel, Joshua Manning, Christopher M. Murtaugh, Margaret Stineman, and Trudy Mallinson

March 2012

Post-Acute Care Payment Reform Demonstration: Final Report Volume 1 of 4

Prepared for

Shannon Flood
Centers for Medicare & Medicaid Services
Center for Medicare & Medicaid Innovation
7205 Windsor Boulevard
Baltimore, MD 21244-1850

Prepared by

Barbara Gage, PhD
Melissa Morley, PhD
Laura Smith, PhD
Melvin J. Ingber, PhD
Anne Deutsch, RN, PhD, CRRN
Tracy Kline, PhD
Jill Dever, PhD
Judith Abbate, PhD
Richard Miller, PhD
Brieanne Lyda-McDonald, MSPH
Cynthia Kelleher, MBA
Danielle Garfinkel, BA
Joshua Manning, BS
RTI International
3040 Cornwallis Road
Research Triangle Park, NC 27709

Christopher M. Murtaugh, PhD
Visiting Nurse Service of New York

Margaret Stineman, MD
University of Pennsylvania

Trudy Mallinson, PhD, OTR/L, NZROT
University of Southern California

RTI Project Number 0209853.005.001

Post-Acute Care Payment Reform Demonstration
Final Report

by Barbara Gage, PhD, Project Director
Melissa Morley, PhD
Laura Smith, PhD
Melvin J. Ingber, PhD
Anne Deutsch, RN, PhD, CRRN
Tracy Kline, PhD
Jill Dever, PhD
Judith Abbate, PhD
Richard Miller, PhD
Brieanne Lyda-McDonald, MSPH
Cynthia Kelleher, MBA
Danielle Garfinkel, BA
Joshua Manning, BS
RTI International
3040 Cornwallis Road
Research Triangle Park, NC 27709
Christopher M. Murtaugh, PhD
Visiting Nurse Service of New York
Margaret Stineman, MD
University of Pennsylvania
Trudy Mallinson, PhD, OTR/L, NZROT
University of Southern California

Federal Project Officer: Shannon Flood

CMS Contract No. HHSM-500-2005-00029I

March 2012

This project was funded by the Centers for Medicare & Medicaid Services under contract no. HHSM-500-2005-00029I. The statements contained in this report are solely those of the authors and do not necessarily reflect the views or policies of the Centers for Medicare & Medicaid Services. RTI assumes responsibility for the accuracy and completeness of the information contained in this report.

RTI International is a trade name of Research Triangle Institute.

CONTENTS SUMMARY

This document represents Volume 1 of 4 of the final report for the Post-Acute Care Payment Reform Demonstration (PAC-PRD). This project was conducted by RTI International under contract with the Centers for Medicare & Medicaid Services. The report has 12 sections, which are divided into four volumes.

- Volume 1: Executive Summary
- Volume 2: Sections 1–4
  - Section 1: Introduction
  - Section 2: Underlying Issues of the PAC-PRD Initiating Legislation
  - Section 3: Developing Standardized Measurement Approaches: The Continuity Assessment Record and Evaluation (CARE)
  - Section 4: Demonstration Methods and Data Collection
- Volume 3: Sections 5–6
  - Section 5: Framework for Analysis
  - Section 6: Factors Associated with Hospital Discharge Destination
- Volume 4: Sections 7–12; References
  - Section 7: Outcomes: Hospital Readmissions
  - Section 8: Outcomes: Functional Status
  - Section 9: Determinants of Resource Intensity: Methods and Analytic Sample Description
  - Section 10: Determinants of Resource Intensity: Lessons from the CART Analysis
  - Section 11: Determinants of Resource Intensity: Multivariate Regression Results
  - Section 12: Conclusions and Review of Findings
  - References

This page intentionally left blank.

CONTENTS

List of Figures

EXECUTIVE SUMMARY

This report is the final report for the Post-Acute Care Payment Reform Demonstration (PAC-PRD). This project was conducted by RTI International under contract with the Centers for Medicare & Medicaid Services (CMS). This report builds on the May 2011 PAC-PRD Report to Congress (RTC) (http://www.cms.gov/reports/downloads/Flood_PACPRD_RTC_CMS_Report_Jan_2012.pdf) and the associated supplemental report (http://www.cms.gov/Reports/Downloads/GAGE_PACPRD_RTC_Supp_Materials_May_2011.pdf). The PAC-PRD was mandated by Congress in the Deficit Reduction Act of 2005 (Public Law 109-171, Sec. 5008) to collect information on PAC populations using a standardized assessment instrument that could uniformly collect data on patients being discharged from acute hospitals to one of four post-acute care (PAC) settings: long-term care hospitals (LTCHs), inpatient rehabilitation facilities (IRFs), skilled nursing facilities (SNFs) and home health agencies (HHAs). The PAC-PRD was also intended to measure patient-specific costs that vary by patient complexity and resource expenditures and that differ from fixed costs associated with the use of specific types of certified providers. Last, the data were also intended to measure outcomes associated with these treatments.

The PAC-PRD was successful in its efforts to develop and apply a consensus-based, uniform approach for measuring medical, functional, and cognitive complexity in Medicare populations and to set national standards for documenting key clinical factors that can be used to monitor the Medicare program. Almost 200 providers, including acute hospitals, LTCHs, IRFs, SNFs, and HHAs, participated nationally to collect data over the 3 years of the demonstration. Feedback from the clinical communities and associations was positive and helpful for refining the items during the development period. The result is an extensive database describing the complexity and costliness of post-acute populations, including both the critically, chronically ill and the healthier Medicare beneficiary who may be admitted to a hospital or use one of the four PAC sites of care.

The PAC-PRD provided information on beneficiaries' medical, functional, and cognitive complexity and the resources used to treat them in each setting. This type of information was needed to better understand the current PAC delivery system, how each type of provider functions within that system, and how provider roles differ according to the availability of alternative options in a local market area. The information also will help in consideration of the implications for improving the consistency of the four Medicare PAC payment policies.

This executive summary summarizes the 12 sections of RTI's four-volume Final Report to CMS on the PAC Payment Reform Demonstration:

- Volume 1: Executive Summary
- Volume 2: Sections 1–4
  - Section 1: Introduction
  - Section 2: Underlying Issues of the PAC-PRD Initiating Legislation

- Section 3: Developing Standardized Measurement Approaches: The Continuity Assessment Record and Evaluation (CARE)
- Section 4: Demonstration Methods and Data Collection

- Volume 3: Sections 5–6
  - Section 5: Framework for Analysis
  - Section 6: Factors Associated with Hospital Discharge Destination
- Volume 4: Sections 7–12; References
  - Section 7: Outcomes: Hospital Readmissions
  - Section 8: Outcomes: Functional Status
  - Section 9: Determinants of Resource Intensity: Methods and Analytic Sample Description
  - Section 10: Determinants of Resource Intensity: Lessons from the CART Analysis
  - Section 11: Determinants of Resource Intensity: Multivariate Regression Results
  - Section 12: Conclusions and Review of Findings
  - References

ES.1 Background

In the Deficit Reduction Act, Congress authorized the PAC-PRD and directed CMS to deliver a report on the results. As indicated by the name of the demonstration, the PAC-PRD was aimed at reforming and harmonizing the disparate methods of paying for services in PAC settings that are, to a degree, either substitutes for one another or complements to each other. In the process, a new patient assessment instrument was to be developed to provide a uniform way of assessing patient needs across settings and to measure the comparability of patients and outcomes.

In the demonstration, patients were assessed at participating LTCHs, IRFs, SNFs, and HHAs, as well as general acute care hospitals. To associate patient characteristics with the resources needed to treat them, data were also collected on the resources used by individual patients. The goal was to provide information that will support the future creation of payment methods that pay appropriately for similar patients irrespective of the setting chosen and provide consistent incentives across the four payment systems.

Almost one in five Medicare beneficiaries are admitted to the hospital each year; among them almost 40 percent will be discharged from the hospital to one of four PAC settings for additional nursing or therapy treatments. In 2008, patients discharged to PAC services tended to go primarily to HHAs (37.4 percent of discharges to PAC) or SNFs (42.2 percent of the PAC users). However, 8.6 percent of those discharged to PAC went to IRFs, and 1.7 percent of those discharged to PAC went to an LTCH. The remaining PAC users received therapy services in either a hospital outpatient department or a therapist's office (see Volume 2, Section 2, for a complete discussion of utilization patterns).

Many of those discharged to PAC used more than one service during their episode of care, particularly those discharged to SNFs and LTCHs. For example, 67 percent of those discharged to SNFs continued on to additional services. Almost a quarter of them were readmitted to the acute hospital (23.1 percent). Another third (32.7 percent) were discharged from the SNF to an HHA. In patients with the acute-SNF-HHA pattern, almost 20 percent (19.9 percent) returned to the acute hospital within 30 days of discharge from the HHA.

LTCH patients were also likely to use multiple types of PAC services. About 74 percent of cases discharged to LTCHs were discharged to additional services after leaving the LTCH, either back to the acute hospital (14.7 percent) or on to an HHA (22.2 percent), IRF (5.7 percent), or SNF (28.5 percent). A substantial share of each of the cases discharged from an LTCH to a third PAC setting were readmitted to the hospital within 30 days of discharge from the PAC service, ranging from 15.9 percent of the LTCH-to-IRF cases returning to the acute and up to 42.8 percent of those discharged from the LTCH to an SNF.

Hospital patients discharged to IRFs were also likely to use multiple PAC services, although the most common third sites of care for this group were in the community. Almost half of the acute-to-IRF cases (47.1 percent) were discharged from the IRF to an HHA; another 17.2 percent were discharged to outpatient or independent therapy. About 16.2 percent were discharged from the IRF to an SNF, and less than 1 percent of these returned to the IRF.

Acute-to-HHA cases typically used only the one service (61.2 percent) unless they were readmitted to the acute hospital within 30 days of discharge (24.3 percent). Of the readmitted cases, 29.8 percent were readmitted to the HHA, and 20.7 percent were discharged instead to an SNF. In examining the home health patterns, it is important to keep in mind that a significant number of the home health population does not come through an acute admission or as part of a post-acute trajectory of care but instead is directly admitted to the HHA from the community. Similarly, those discharged from the hospital to outpatient therapy (6.7 percent) or other independent therapists (3.4 percent) typically used only that one post-hospital service.

In general, the four PAC settings are assumed to differ in the type and intensity of services provided, effectively providing a "continuum of care." But these providers' services are not mutually exclusive; each of the three inpatient PAC settings (LTCHs, IRFs, and SNFs) provide 24-hour nursing, and all four settings provide physical, occupational, and speech pathology services to some extent.

Currently, Medicare uses a different prospective payment system (PPS) for each of the PAC providers, each with its own case-mix groups, payment units, associated payment rates, and incentive structures. However, many conditions may be treatable in more than one of these PAC settings, making these settings potential substitutes for treating the same type of patient. Past research has shown that patients treated for the same condition in an acute hospital may be discharged to different types of PAC settings for subsequent treatment, depending on the availability of PAC options in the local market and other factors not measurable in the Medicare claims data (Gage et al., 2009; Gage, 1999).

This situation prompted the need for consistent measurement approaches in order to determine the patient characteristics that influence resource needs and treatment costs, to evaluate patient complexity in a consistent manner between settings, and to measure associated resources and outcomes in each setting and across settings.

Currently the PAC payment systems differ in how they measure patient severity and form case-mix groups for payment and quality reporting purposes. Three of the PPS case-mix groups (IRF, SNF, and HHA) are based on assessment data that measure patient complexity factors not found in the claims data. Although the concepts are similar in each PPS, the exact items used to measure patient complexity differ across the three systems. The fourth PPS (LTCH) relies entirely on claims data for measuring severity, limiting measures largely to diagnoses and procedures data.

Because each PAC PPS uses different case-mix measurement items, it has been difficult to compare the populations admitted to each site and the costs and outcomes associated with treatment in the four PAC sites. These issues are further complicated by the different episode patterns, which may include several types of PAC service use during an episode of care and, depending on local availability, may use alternative types of settings for similar services. Understanding the factors that drive these different utilization patterns is necessary to ensure that appropriate payment incentives are aligned for each of the PAC providers. While the settings may be paid individually, together they represent a beneficiary's complete episode of care.

ES.2 The CARE Tool Items

The Medicare program currently mandates that IRFs, nursing facilities (including SNFs), and HHAs submit assessment data on the beneficiary's medical, functional, and cognitive status. The information collected through these assessments is used by CMS to calculate payment groups, generate quality measures, and monitor regulatory compliance, and by many states for Medicaid payment and quality monitoring. These assessment instruments are usually referred to by their acronyms, IRF-PAI (Inpatient Rehabilitation Facility Patient Assessment Instrument), MDS (Minimum Data Set), and OASIS (Outcome and Assessment Information Set), respectively. LTCHs and acute hospitals do not have standardized assessments, although they all use variations on these measures to conduct assessments at intake and throughout the hospital stay. However, the measures used in general acute hospitals and LTCHs are not standardized across hospitals and, for certain items, the data may be found only in medical notes. The current assessment systems differ in other ways as well, even among the three federally mandated assessments. The MDS, OASIS, and IRF-PAI have incompatible data formats; thus, it is difficult to share data electronically across levels of care. Within settings that have integrated data systems across different levels of care, the three federally mandated tools are either excluded or have to be incorporated by the software vendors into the existing system. Further, each tool uses different assessment windows, resulting in the patients being assessed at different times during their treatment period. Patients in the LTCH are typically assessed at admission and throughout the stay; IRF admissions data reflect the first 72 hours of the stay; SNFs submit data reflecting the first 5 days of the admission; and HHAs submit initial assessment data related to the first visit, which is tied to the physician's ordered start date or within the first 48 hours of referral or return home. HHA staff have 5 days to complete the comprehensive assessment. These differences make it difficult to compare severity, outcomes, and cost across providers. The three mandated assessments all measure similar concepts, but they use different clinical items, timeframes for data collection, and measurement scales. A common assessment tool that could be applied across settings including acute care hospitals and LTCHs is needed to permit comparison of populations within and across PAC settings and to evaluate transfers or outcomes of similar populations associated with different settings.

To address this need and to comply with the Congressional directive, CMS developed a uniform assessment instrument to measure the range of patients seen across the participating provider types: the Continuity Assessment Record and Evaluation (CARE) tool. As mandated by Congress, the CARE tool was designed to collect data on patients' medical, functional, and cognitive status at admission and discharge from each PAC setting and at discharge from general hospitals. The CARE items are based on the current state of knowledge in assessing patient acuity and outcomes measures and experience in what has been found to be important in the current payment systems, and they represent standardized versions of items being collected in each setting. For the time-sensitive data, CARE established standard assessment windows (timeframes) of the first 2 days following admission and the last 2 days of a stay prior to discharge. This created uniform assessment windows across the different settings to examine patient severity at admission and at discharge. The development of CARE is described below and in greater detail in Volume 2, Section 3, of the Final Report.

ES.2.1 Guiding Principles of CARE Tool Development

The CARE tool's development was based on certain guiding principles. As required in the Deficit Reduction Act, the CARE tool needed to meet certain goals:

- The CARE tool should be designed to collect standardized information at discharge from acute hospitals and at admission and discharge from the four PAC providers: LTCHs, IRFs, SNFs, and HHAs.

- The CARE tool items should inform payment policy discussions by including measures of the needs and the clinical characteristics of the patient that are predictive of resource intensity needs.

- The CARE tool items should inform the evaluation of treatment outcomes by including patient-specific factors that measure outcomes and incorporate the appropriate risk adjustment factors. Outcomes should include but not be limited to measures of functional status.

- The CARE tool items should document clinical factors associated with patient discharge placement decisions to allow the clinicians treating the patients to make appropriate discharge placement decisions.

- The CARE tool should be appropriate for collecting standardized patient assessment information as a patient is transferred from one setting to another and, by standardizing how information is collected, foster high-quality, seamless care transitions.

Individual item selection was based on several overriding principles:

- Sensitivity to data collection burden. Selected concepts and items were restricted to those that were typically already in use for payment or quality reporting purposes or would improve these efforts. Further, only a small subset of items are designated as core items collected on all patients; the larger subset of items are selectively used to define severity of a condition when a condition is present. Few items apply to all patients.

- Consideration of the reliability and validity of items. Items included in the Federal set needed to be reliable and valid measures of the concepts they were intended to measure. Extensive testing of the reliability and validity of the items was needed to consider whether the standardized version in the CARE tool was as reliable and valid as the item in the original tool (MDS, OASIS, or IRF-PAI).

- Breadth of application to minimize floor and ceiling effects. Certain items in the existing tools were limited in their ability to measure acuity for the very sickest and the very healthiest patients (floor and ceiling effects) and thus in their ability to explain variation across patients having a broad range of severity within the measured clinical characteristics found in the PAC populations. These items were selected to reduce those limitations in the current tools.

- Minimization of "gameability" or incentives that might encourage provider behavior that is inconsistent with best practices for patient outcomes and care quality. Different items were tested to identify patient factors that could be substituted for resource measures in the current system. Factors needed to be reliable, objective, and not discretionary in nature.

ES.2.2 CARE Item Approach

The CARE item set was organized to minimize provider burden. Two types of items are included in the set: (1) a small set of core items that provide basic information on patient severity and screen for complicating conditions, and (2) supplemental items that measure the severity of a condition once identified by a screener item. The majority of items are supplemental and are used to measure severity of a condition only if a condition is present. Hence, most factors are not assessed on every patient, but those items that are relevant are collected in a standard way. Estimated burden ranged from a 30-minute assessment completion time for the healthier patients to 60 minutes in LTCHs or SNFs, where patients may be more complicated medically or functionally or have greater cognitive complications. These average times of completion reflect experience with the tool, following training on the appropriate measurement methods, and are consistent with current intake assessment times.[1]

The four clinical domains included in the CARE item set are as follows:

- **Medical Status/Clinical Complexity.** These items measure patient medical status and include factors defining complexity in terms of medical diagnoses, resource use such as procedures or major treatments received during stay (e.g., ventilator weaning, hemodialysis), medications, skin integrity (number and size of pressure ulcers and locations and presence of other wounds), and physiologic factors (e.g., vital signs, laboratory results, blood gases, pulmonary function).

- **Functional Status.** These items include screening items on impairments (e.g., bladder, bowel, swallowing, vision, hearing, weight-bearing, grip strength, respiratory status, and endurance), as well as measures of self-care, mobility, and safety-related functions (medication management, phone management) and other items relevant to less impaired populations.

- **Cognitive Status.** These items target memory/recall ability; delirium/confusion (some of which may be short term related to current medications or longer term, which may complicate rehabilitation therapy); behavioral symptoms, including those that are self-injurious (pulling IV lines) or directed toward others; signs of depression or sadness; and presence of pain, which may affect patients' engagement and outcomes.

- **Social Support Factors**. These items target social support issues, including information on structural barriers, living situations, caregiver availability, and the need for assistance, as well as issues related to discharge complications.

Together, these four domains provide a comprehensive overview of a patient. For healthier patients, fewer items are relevant. For the more complex patients, the CARE items offer standardized versions of information already typically collected on those types of patients.

ES.2.3 Stakeholder Input

The conceptual domains and specific items were selected by the major stakeholders and subject matter experts including clinicians, policymakers, providers, and national professional and provider associations. Some of the participating associations included American Health Care Association, American Hospital Association, Acute Long Term Hospital Association, American Medical Rehabilitation Providers Association, Commission on the Accreditation of Rehabilitation Facilities, The Joint Commission, Leading Age (formerly American Association of Homes and Services for the Aging), National Association for Home Care, the National Association of Long Term Hospitals, and the Visiting Nurse Association of America. Additional

[1] These items are intended to replace nonuniform versions of the items already used and would not add any time relative to the current items. They added time in the demonstration because providers needed to continue collecting the mandated version for reimbursement while also collecting the test version during the study period.

input was provided throughout the process by several clinical communities, including the National Pressure Ulcer Advisory Panel, the Association of Rehabilitation Nurses, the American Academy of Physical Medicine and Rehabilitation, and others. These provider associations and the clinical and measurement experts provided valuable input regarding the types of concepts to distinguish severity and the items that best measured those concepts across all settings.

Stakeholder and other public comments were incorporated in multiple stages and through multiple avenues including open door forums, technical expert panels, presentations to provider groups and other interested parties, input submitted through a project website, and a widely publicized e-mail address.

The CARE data set includes elements covering administrative information, premorbid health status information, current medical status, measures of cognitive status, pain, impairments, functional status, and discharge information. Though much of this type of information was already collected by the existing instruments, the specific items used often varied. The CARE items are uniform across the settings, including those settings that have not used a mandated assessment but collect this type of data as part of their current intake and assessment process.

ES.2.4 Item Validity and Reliability Tests

The CARE tool and the items included in the CARE tool were extensively evaluated and tested during the development process and in specific reliability tests during the demonstration. In the development phase, two sets of pilot tests were conducted in the Chicago area. Although the sample sizes were small in the pilot tests, they provided important preliminary information on the feasibility of using each item in the different treatment settings before testing the items in a national demonstration. The validity and reliability of the CARE items' use in each setting was evaluated as part of the demonstration.

Validity and reliability were tested through two methodologies. First, practicing clinicians were asked for feedback on the items' use with different types of patients in their respective settings. Second, two types of formal reliability tests were conducted. The first used a traditional inter-rater reliability study approach to focus on the reliability of the standardized items when applied to populations in settings other than those for whom the items were originally validated. The second type of test, where assessors in different settings rated uniform "hypothetical" patients, examined the degree of agreement when items were used by different disciplines in different settings. In addition, the validity of CARE items was assessed relative to existing items in the legacy tools (MDS, OASIS, and IRF-PAI), and the parsimony of the measurement approach was evaluated.

Overall, the results showed very good agreement on most items. The reliability results were consistent with those achieved in earlier efforts testing the nonstandardized items and suggest they can be used to replace the items in the current legacy tools. Across all 146 items tested, only 17 percent had a rating lower than 0.60, including both the unweighted and weighted kappas and in samples with and without missing values included. In summary, the key findings include the following:

- Most of the standardized CARE items performed reliably across settings. All five settings were able to collect information in a reliable, consistent, and comprehensive manner for their Medicare populations.

- Participant feedback on CARE items was generally positive. Clinicians in all five settings appreciated the use of standard items for measuring pressure ulcers and other medical factors that affect staffing intensity. Therapists consistently commented that the CARE items were easy to use and provided greater specificity for measuring severity and change in function than the items that had been in the MDS 2.0 and OASIS-B in use at the time of the demonstration. They also commented positively about the coding approach of determining whether a patient could do at least half the task or not, and if they could, whether they could safely leave the patient to complete the task without supervision. The LTCH staff appreciated being able to note small changes from complete dependence to being able to complete a task with much assistance (over half the task was completed by the helper), particularly for the most impaired populations.

- Reliability testing for CARE showed positive results that are consistent with reliability standards used for previous CMS mandated patient assessment instruments, suggesting that these items can be used in each setting and be reliable enough for payment and quality monitoring purposes.

- Overall, the inter-rater reliability results showed very good agreement on most items. These results suggest that most of the standardized versions of the assessment items have strong reliability within and across settings. Differences across settings were present, but each setting still had acceptable levels of reliability within settings, suggesting that these items could be used to measure a patient's progress in a standardized way across an episode of care.

- Items with poorer agreement among any of the samples (less than 0.60) tended to be items with fewer responses (e.g., items where the response code was "other" or "tube feeding" and "comatose," for which few cases were included). A few items with reasonable sample sizes appeared to be less reliable, such as certain components of the swallowing item ("complaints of difficulty or pain when swallowing," "holding food or liquid," and "loss of liquid when swallowing"). These lower reliability ratings were offset in the swallowing item by less discretionary components, such as "no intake by mouth" (NPO; 0.97) and "no impairments" (0.84). Other poor-scoring items included "walking 150 feet," "light shopping," and "laundry." These items were not used in the analytic models.

ES.3 Data Collection

Data collection required consideration of the patient populations, the types of care they received, the settings in which they received it, and the variation in practice patterns that occur in the Medicare program (Section 4). Market areas and providers were selected to account for the following factors: (1) variation in the supply of providers of different types; (2) geographic variation; and (3) beneficiary/patient representativeness. Data were collected from 206 providers across the country for a total of 53,952 assessments included in these analyses.

Two types of data were collected from the participating providers. All providers, including both acute hospitals and PAC providers, collected the CARE standardized assessment item set discussed above to provide data on patient complexity. To provide data on the resources used to treat patients of different types, PAC providers also participated in a set of staff-time studies. These involved submitting cost and resource use (CRU) data, which included staff time measures for treating a subset of the assessed beneficiaries in each setting. Participating provider units collected data on staff time spent with each Medicare patient during three 2-week-long data collection windows within the 9-month CARE collection period in each facility. The HHA data were collected as visit time by licensure type and were also assessed through claims information. In the supplemental phase of the data collection, both the post-acute and the acute providers submitted CRU data.

ES.4 Analytic Framework

Section 5 in Volume 3 of the report presents the conceptual framework for assessing patient complexity. A comprehensive framework must allow for inclusion of multiple factors, ranging from those with the widest applicability to those with the narrowest scope. As such, this framework, which is intended to explain variations in costliness and outcomes, needed to include all three types of health status: medical, functional, and cognitive.

These three domains—medical, functional, and cognitive—are currently collected in at least one of the four PAC payment systems as factors that predict variation in resource intensity. Each of these components of health status is important for defining case-mix criteria and may affect the patient outcomes independently or by interacting with other patient characteristics. The proposed classification scheme builds on the current PAC case-mix systems to use standardized versions of items already in each respective PPS.

This classification framework builds on the logic of the current Medicare classification systems, which vary in their recognition of medical, functional, and cognitive factors in these populations. For example, the LTCH PPS uses Medicare Severity-Diagnostic Related Groups (MS-DRG) to classify patients based on medical complexity. The MS-DRG system uses ICD-9 codes to define the primary condition, whether they were medical or surgical in nature, and assigns a severity of illness level based on complicating comorbidities, as all of those factors affect the relative complexity or costliness of patients at that level of illness. Although cognitive status may be impaired, it is assumed to consistently affect the costliness of nursing care in each diagnostic group and is not measured separately. If the effect of the cognitive condition varies within a case-mix group, it is directly measured as a complicating condition by including an ICD-9 code for the condition in the severity adjustment (e.g., dementia as a complicating severity factor within a DRG). Functional impairments are not used in classifying LTCH patient complexity, although many LTCHs provide specialized therapy services in addition to the medical treatments, and these effects may be variable within MS-DRG groups. Given this, separate recognition of function may be valuable for improving the predictive power of LTCH case-mix classification systems.

The IRF payment policies use medical, functional, and, for some cases, cognitive factors to classify a patient's complexity. Primary reason for treatment is defined by ICD-9 codes that specify the etiologic or underlying medical condition. In this system, the etiologic or primary reason for treatment is used to classify the case, and the comorbidities are used to adjust payments. Functional status, cognitive status, and age are also taken into account.

Similarly, SNF payment policies also use medical, functional, and cognitive factors in the resource utilization groups (RUGs) case-mix system. The primary reason for treatment is less important than the total constellation of medical factors in this system. SNF medical conditions are identified by an indicator of whether a patient has certain medical conditions without distinguishing between primary and secondary diagnoses. Medical complexity is further defined by the presence of other medical factors, such as pressure ulcers and the need for ventilators, to name a few payment factors. Function and cognition are also taken into account.

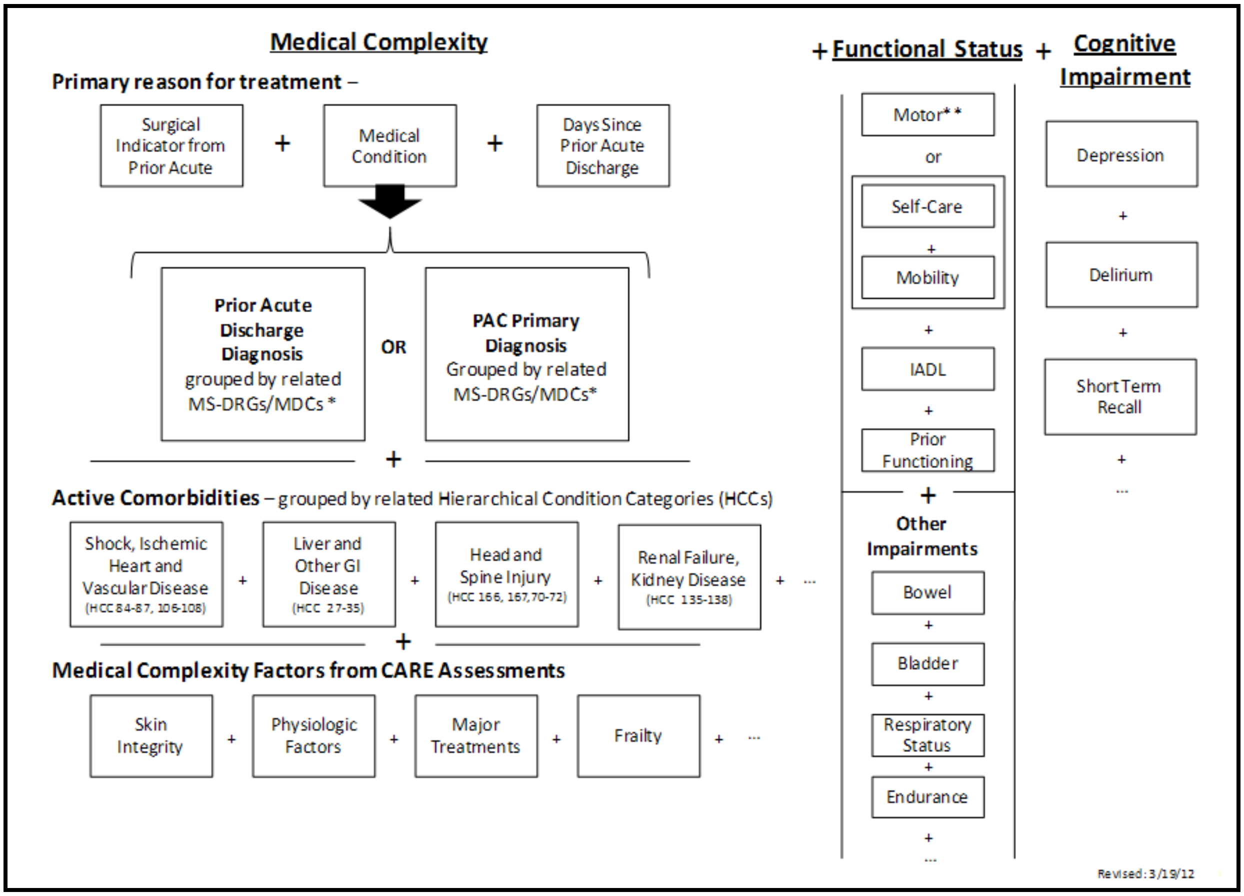
HHA payment policies also use medical, functional, and cognitive factors, but HHAs must report both primary reason and comorbid conditions using ICD-9 codes. HHA case-mix adjustment includes large grouping of medical conditions, some based on the primary diagnosis only and others based on all diagnoses listed. Like the SNF policies, medical conditions are further identified by additional medical complications, such as pressure ulcers and other factors. Both HHA and SNF coding systems may use a procedure (or a V code) as the primary reason for admission.

As noted above, the IRF PPS, HHA PPS, and SNF PPS all use medical, functional, and cognitive status to determine case-mix groups. Each of the three systems measure some mix of function items, including 18 physical and cognitive items in the IRF, five activities of daily living (ADL) or mobility items in the HHA, and four ADL items in the SNF. However, both the HHA PPS and the SNF PPS also include a resource utilization measure (number of therapy visits in HHA and count of therapy minutes in SNF). Although these additional measures produce strong results, they are based on resource use rather than patient severity, a less desirable approach for predicting costs.

Our approach assumes that each of these three domains—medical, functional, and cognitive status—may predict resource needs because they define severity of illness, difficulty of treatment, need for intervention, and the expected volume and types of routine or therapy resource intensity. The measures used in these analyses are based on the patient characteristics and avoid the use of utilization measures to predict resource intensity.

The analyses presented in this report test the extent to which each of the three domains is important in each setting and identify the best measures of each concept by testing their potential contribution to explaining resource intensity and treatment outcomes. **Figure ES-1** shows the classification schema underlying our approach, which is described in detail in Volume 3, Section 5, of the report.

Figure ES-1
CARE Case Mix Classification Schema



*A modified MS-DRG/MDC system was used in the analysis (e.g., the neurologic major diagnostic category (MDC 01) is subdivided into neurologic, stroke (MS-DRGs: 020, 021,022,061-066), neurologic, surgical (MS-DRGs: 024-042), and neurologic, medical (MS-DRGs: 052 -060, 067-103)). Similarly, the Hierarchical Condition Categories (HCC) classification was modified slightly for use in this project.

** The motor scale combines the self-care and mobility scales, which are listed separately in this section as well.

NOTE: Where the complete list of factors under each category is not presented in this chart, this is indicated by the notation: '+ …'.

ES.5 Analysis of Factors Associated With First Sites of PAC

One of the key goals in this demonstration is to better understand the types of patients treated in each of the four PAC settings, including LTCH, IRF, SNF, and HHA. The use of a standardized assessment tool allows examination of the populations admitted to each setting in greater detail than possible using claims data. It also allows populations to be compared across settings to identify factors that distinguish admissions to each setting, as well as identify overlapping characteristics that may be useful for understanding whether the same patient is treated in more than one setting. Section 6 in Volume 3 examines how patient complexity (medical, functional, and cognitive) factors are associated with hospital discharge home or to a PAC setting. The focus is on the use of a Medicare-covered PAC service during the first 30 days from acute discharge, and, if there was use, on the factors associated with the type of first PAC site used. It is important to note that these analyses are based on current practice patterns and do not necessarily reflect an "ideal" system of care or PAC decision.

The analyses presented in this section are important for understanding the extent to which types of patients treated in each setting overlap or are distinguishable. These issues are important because Medicare uses a different payment system with different payment units, case-mix groups, and payment amounts for each type of provider. Hence, Medicare may be paying different amounts for similar types of patients who may be treated in more than one setting. Further, outcomes may differ depending on the type of PAC setting used. Understanding differences in the complexity of post-acute patients admitted to each type of setting and the outcomes associated with their treatment will be important for considering future payment reform. These issues are complicated by variations in the availability of the more specialized PAC settings, such as inpatient rehabilitation hospitals and long-term care hospitals compared to the widely available skilled nursing facilities and home health agencies. Understanding whether similar populations can effectively be treated in more than one setting, and the availability of those services, is important for a variety of policy-related issues.

This analysis builds on much of RTI's past work using Medicare claims to predict the discharge destinations of acute hospital patients. The claims data are useful as a first stage in measuring medical complexity, but they fail to measure more specific areas of medical, functional, and cognitive health status complexity. The standardized CARE items provide much more additional detailed information on these areas.

In examining the issue of discharge destination, there may be a number of different ways to consider factors associated with PAC use. Most prior analyses have been limited by data availability and rely on the patient factors available in the claims data, or, if studying one setting of care, the assessment data such as the MDS, IRF-PAI, or OASIS data associated with that type of setting (Gage et al., 2008; Wolfe and Meadow, 2008; Gage et al., 2005; Gage, 1999; Liu, Wissoker, and Rimes, 1998; Lee et al., 1997; Kane et al., 1996; Kramer, Shaughnessy, and Pettigrew, 1985). The studies have varied in whether they looked specifically at the use of certain types of PAC, such as the choice of HHA use versus discharge home without services (Kenney and Moon, 1994), or whether they looked across the range of PAC providers either in comparisons of dyads (SNF versus IRF choice) or more generically (PAC service versus no PAC service) or in multisetting models (predicting no service use versus SNF use versus IRF use (Gage et al., 2009; McCall et al., 2001; Lee et al., 1997).

ES.5.1 Sample and Methods

This analysis presents four different approaches to examining factors associated with PAC use following discharge from an acute hospital stay. Each of these models controls for medical, functional, and cognitive status as well as market, which vary in characteristics, such as the availability of PAC options. First, a logistic regression is used to predict the probability of any post-acute use within 30 days of hospital discharge (ANYPAC). Post-acute use in this model is broadly defined as receiving services in one of the four PAC settings examined in this project (LTCH, IRF, SNF, or HHA) or any Part B therapy, or subsequent hospitalizations. The contrast is discharge home with no services in the 30-day window. Second, a multinomial logistic regression (FIRSTPAC) is presented to compare the relative odds of being discharged home without Medicare-covered inpatient or home health services relative to being discharged to (1) home health care, (2) skilled nursing facility, (3) inpatient rehabilitation facility, or (4) long-term care hospital.[2] Third, two additional logistic regressions examine the relative characteristics differentiating between hospital discharges to (1) skilled nursing facilities versus inpatient rehabilitation facilities (SNF/IRF) and (2) skilled nursing facilities versus home health agencies (SNF/HHA).

ES.5.2 Sample

The sample for these analyses is based on 13,554 cases with a CARE assessment at the time of transfer from the inpatient prospective payment system (IPPS) to the PAC setting. CARE at time of transfer could originate from two sources: (1) the CARE assessment for patients at the time of discharge from acute hospitals (n = 4,412) or (2) the CARE assessment at the time of admission to the PAC setting (n = 9,020). The PAC admission sample was restricted to admissions occurring within 2 days of hospital discharge dates.[3] The analysis focused on the first discharge destination after a hospital stay and thus excluded home health cases that were not PAC and cases that were secondary PAC admissions.

The providers targeted for inclusion were selected to examine Medicare PAC populations. Hospital units included in this study were selected for treating Medicare populations with the types of diagnosis that frequently are discharged to PAC (i.e., stroke, cardiac, chronic obstructive pulmonary disease [COPD], to name a few). As noted in Section 4, this sample represents a higher proportion of PAC users than the national sample, but it allows for a larger number of cases that had the potential to use PAC. This approach reflects the analytic focus on addressing payment equities across PAC systems. Hence, the analyses presented here are primarily useful for understanding the types of cases treated in each setting or going home without services, but they do not reflect the entire Medicare population likely to go home without PAC. Second, the markets were selected for having higher or lower options for PAC. The high PAC markets have IRFs or LTCHs in addition to SNFs and HHAs, whereas the low PAC markets reflect the absence of these more specialized services. These factors may also influence discharge destinations.

ES.5.3 Methods

Four analytic models were tested in Volume 3, Section 6. Each of the models is measuring the odds of admission to a PAC or other service within 30 days of discharge from the acute hospital. If more than one service is used during that window, only the first type of service is considered for the discharge destination. Service use is based on the first Medicare claim

[2] The FIRSTPAC model excludes claims from the "other" settings, including Part B therapy, Federal hospitals, and inpatient psychiatric hospitals.

[3] Service utilization sequences were based on Medicare claims for the cases in our CARE sample. Discharges home without services were defined as cases having no Medicare claim (acute, LTCH, IRF, SNF, HHA, or Part B therapy) within 30 days following hospital discharge.

within 30 days of discharge from the short-stay acute hospital. We tested several different discharge destinations as the dependent variables.

- **Any PAC model:** This outcome was defined as a yes/no indicator of whether the beneficiary had a Medicare claim (LTCH, IRF, SNF, HHA, Part B therapy, hospitalization) within 30 days following discharge from the hospital. Note that cases with a zero day transfer from an acute hospital to another acute hospital, as noted above, were excluded from the analysis.

- **First PAC model:** This outcome was defined as one of five outcomes based on the first site of PAC used within 30 days following discharge. The multinomial model predicts the odds of using one of the following settings relative to not having a claim for any of the following services. Cases discharged to "other" were excluded from this analysis. The settings were defined as follows:

  - Long-term care hospital (LTCH)

  - Inpatient rehabilitation facility or hospital unit (IRF)

  - Skilled nursing facility (SNF/transitional care unit [TCU])

  - Home health agency (HHA)

- **IRF/SNF outcomes:** This outcome selects only IRF and SNF admissions to examine more closely the factors associated with discharge to either of these services. This outcome was defined as having either an SNF or an IRF claim within 30 days following discharge from the hospital. The SNF group was the referent category.

- **SNF/HHA outcomes:** This outcome selects only SNF and HHA admissions to examine more closely the factors associated with discharge to either of these services. This outcome was defined as having either an SNF or an HHA claim within 30 days following discharge from the hospital. The SNF group was the referent category.

The goal of these analyses is to examine the medical, functional, and cognitive factors associated with Medicare service use following hospital discharge. The independent variables used in this analysis include demographic, medical, and functional characteristics; mood and cognition; and indicators of premorbid functional status and premorbid living arrangements as noted in the classification schema presented in Volume 3, Section 5, of this report.

The analysis was the first large-scale analysis using standardized items that could allow comparisons of the populations being admitted from an acute stay to each setting. By using the same approaches to specify precipitating medical conditions and existing comorbidities, and the same measures of pressure ulcers, history of falls, functional status and impairments, and cognitive status, one could finally consider whether the same types of patients are treated in more than one setting and start to discuss differences in clinical complexity that may exist within and across PAC settings. These analyses do not attempt to answer the question of where patients should go, but instead examine the existing patterns of care given the regulations and incentives currently in the marketplace.

ES.5.4 Results

The results showed, on average, variations in referral patterns in the markets examined. The types of patients treated in each of the four settings had overlapping characteristics, although the odds of using each type of service may have differed by individual characteristics.

PAC users tended to be either the younger disabled or the older Medicare beneficiaries. Their primary diagnosis in the acute hospital ranged across both medical and rehabilitation types of conditions. Comorbid conditions were common across the populations using PAC. Having symptoms of depression was associated with a higher odds of using PAC as was having a history of falls or some type of physical or communication impairment. Living alone in the community was also associated with higher odds of being discharged to PAC.

Although the complexity of patients using each PAC setting tended to differ across settings, the results suggest that the populations using PAC also appeared to overlap in the types of conditions and impairments being treated in each level of care. Notably, the results showed that medical cases were more likely to be discharged to HHAs, SNFs, and LTCHs, whereas post-surgical cases typically needing physical rehabilitation tended to be discharged to IRFs, SNFs, and HHAs. Medical factors, such as primary diagnosis in the acute hospital, were important but not sufficient for predicting subsequent PAC use. Comorbidities played an important role in identifying the difference in the potential complexity of cases treated in each setting. For example, the odds were greatest for the LTCH setting when more medical comorbidities were present. However, when the comorbidities were the type that required therapy services, such as orthopedic/musculoskeletal conditions and neurological conditions, patients had higher odds of IRF or SNF use. Similarly, cases with higher medical resource needs, such as being discharged on a ventilator, requiring hemodialysis, or being discharged with no food intake by mouth (NPO), were all associated with greater odds of being discharged to an LTCH. Interestingly, after controlling for the other factors in the model, having had an intensive care unit (ICU) stay longer than 7 days did not increase the odds of going to an LTCH.

Functional status was also an important factor in explaining site of care. Although IRF patients frequently have a history of falls, the models suggest that after controlling for the other patient characteristics, patients with a history of falls have no higher likelihood of being discharged to an IRF. However, falls history is significantly associated with higher odds of being discharged to an SNF, all other patient characteristics being equal.

The relationship between the self-care and mobility score at time of transfer and discharge destination was curvilinear in nature. In other words, although the SNF and IRF have significantly higher odds of taking patients with higher self-care scales, the square term is negative, suggesting a lower likelihood of patients being discharged to these settings once the self-care score is too high. Similar results are shown with mobility scores, although the two settings with the higher odds of accepting patients with higher scores at admission are HHAs and IRFs, but again, these scales reach a point where the patient is significantly less likely to be admitted to these settings, although the difference is very small. And as with the medical characteristics, these factors are significant in more than one setting, underscoring the overlap in patients admitted to the different sites of care.

Cognitive impairments were also significantly associated with PAC use. Depression was associated with higher odds of using all four PAC settings, although HHAs to a lesser degree.

Two of the models presented, the SNF/IRF and the SNF/HHA, allowed better understanding of the characteristics differentiating treatment between these settings. It was notable that neurological patients had significantly higher odds of being discharged to an IRF than an SNF when the sample was restricted to the two groups, but many of the other diagnosis and comorbid factors remained similar to the multinomial model. However, this is a relative finding and not suggestive that these cases are not treated in SNFs.

The role of HHAs in treating some of the more chronic populations was also notable. After controlling for primary diagnosis and comorbidities, the cases with severe respiratory status impairments and those with limited endurance (could endure with support or rest) had higher odds of being discharged to HHAs than SNFs. However, cases with a history of falls had higher odds of being discharged to an SNF than an HHA, perhaps related to the concern over patient safety when discharging them to the home environment.

Together, these results present a picture of the constellation of factors associated with patients in these settings. Medicare patients are complex. Unlike younger, nondisabled populations, Medicare beneficiaries tend to have multiple factors affecting their general health status. These analyses were useful for empirically identifying some of the overlapping characteristics and beginning to consider the ways in which PAC populations or subpopulations may differ. The findings showed that patients with these types of medical, functional, and cognitive factors generally had a higher probability of using PAC than being discharged home without further services. Although the magnitude may vary by setting, these findings underscore that PAC settings do treat overlapping populations. Understanding whether treatment outcomes and resource intensity associated with treating these cases differs across the PAC settings is needed to consider the appropriate approach for payment reform.

ES.6 Outcomes

Sections 7 and 8 in Volume 4 of the report present the information related to the analytic approach and results associated with selected outcomes of interest.

ES.6.1 Analytic Approach

The outcomes analyses were important for understanding whether different types of PAC settings achieved different outcomes after controlling for patient characteristics. Three outcomes were examined: (1) change in self-care functioning from admission to discharge, (2) change in mobility functioning from admission to discharge, and (3) readmission to the hospital within 30 days.

Regression models were used that included patient characteristics at admission to the PAC setting and setting indicators. The size and significance of the coefficients on the setting indicators were interpreted as measures of the effect of setting on the outcome after controlling for patient acuity. The readmission outcome was a simple yes/no variable for each patient indicating readmission to an acute hospital for any cause within 30 days of hospital discharge. The function variables measured change in the function scales from admission to discharge. These admission and discharge function scales ranged from 0 to 100 and were created by combining a number of related self-care or mobility function items from the CARE tool. The items were combined into a Rasch measure, which incorporates patient ability and the difficulty of each function item into how the scale is created.

In attempting to interpret the results of the outcomes analysis, several issues should be kept in mind. First, it should be noted that these analyses focus on outcomes per PAC stay and not on differences in daily effects or episode of care effects. The SNF stay was on average twice as long as the IRF admission, while the HHA effects are related to a complete HHA admission, regardless of the number of 60-day episodes. Second, in controlling for patient acuity, the models focused on patient acuity factors measured at admission to the PAC setting. Many factors such as patient involvement in care and family engagement were not included in the models but could be correlated with both the likelihood of treatment in a particular setting and the likelihood of a favorable outcome.

ES.6.2 Results

ES.6.2.1 Changes in Self-Care Function

Across the whole sample and the condition-specific samples, HHAs admitted patients with the highest mean unadjusted self-care measures (overall: 59.9, musculoskeletal: 58.5, nervous system: 55.5), and LTCH patients had the lowest (overall: 33.9, musculoskeletal: 41.8, nervous system: 33.1), suggesting that, on average, the patients admitted to HHAs were the least impaired in self-care and LTCH admissions were the most impaired. Cases admitted to IRFs were slightly more impaired than those admitted to SNFs (43.6 compared with 45.4 at admission, respectively). This was true in both the musculoskeletal and nervous system subpopulations also. At the same time, it is important to note that all four settings treated patients with a range of functional ability, and no one setting exclusively treated a particular type of patient.

Overall, the mean unadjusted change in self-care function was 12.4, with a standard deviation of 13.8 units. In looking at the unadjusted data, IRF patients had the greatest increase in self-care overall (15.5 units) and within each of the subpopulations examined (17.4 units in the musculoskeletal and 13.8 units in the nervous system patients). SNF patients achieved the second highest change scores in the overall patients (12.4 units improvement) and in the musculoskeletal patients (15.5 units improvement). Within the nervous system populations, SNFs achieved 10.1 units improvement. HHAs achieved improvements in self-care that were roughly comparable to SNFs in the overall population (10.0) and in the musculoskeletal population (14.6). HHAs had slightly lower improvement rates in the nervous system group (7.8). Unadjusted LTCH rates for the diagnosis subpopulations tended to be lower but reflect a smaller sample size.

After adjusting for patient characteristics, we found that IRFs and HHAs had a significantly greater improvement on self-care outcomes than SNFs, with some variation in results associated with different diagnosis groups. Across all conditions, IRFs achieved a 30 percent better self-care status at discharge than SNF patients, after controlling for patient acuity characteristics at PAC admission. HHAs had a 32 percent better self-care outcome than SNFs, after controlling for patient case-mix differences. These may be related to unmeasured factors such as patient levels of engagement, differences in family involvement, and length of stay in these settings relative to a SNF. At this point in the analysis, caution should be taken in assigning causation to these results.

The impact of setting after controlling for multivariate effects differed by diagnosis. For musculoskeletal cases, HHAs had 35 percent better gain in self-care outcomes than SNFs; IRFs and LTCHs had no significantly different self-care outcomes than SNFs. For patients with nervous system disorders, including stroke cases, IRFs achieved 32 percent better functional improvement in self-care than SNF patients at discharge, while HHA and LTCH patients were not statistically different from SNFs.

In summary, key findings related to the prediction of change in self-care functional ability include the following:

- After controlling for the patient acuity measures, provider type is a statistically significant predictor in the models of change in self-care functional ability from admission to discharge. Both IRF and HHA stays were associated with a positive impact on improving self-care functional ability from admission to discharge relative to SNFs after controlling for the patient acuity measures.

- The relatively significant positive impact of the IRF and HHA settings held for some but not all diagnosis groups examined.

- The self-care change results are preliminary and it is not possible to ascribe causation to specific interventions. The models control for many patient acuity factors but do not attempt to examine the impact of many psychological and social factors that may vary systematically between settings.

ES.6.2.2 Changes in Mobility Function

Across the whole sample and the condition-specific samples, HHAs had the highest unadjusted mean admission mobility measures (overall: 59.9, musculoskeletal: 57.3, nervous system: 54.0), and LTCHs had the lowest (overall: 33.5, musculoskeletal: 37.0, nervous system: 33.7), suggesting that, like in self-care, while substantial areas of overlap exist between settings, the patients that were least impaired in mobility were treated in HHAs and the most impaired in LTCHs.

The mean change in mobility for the overall sample was 14.6, with a standard deviation of 14.6 units. IRFs and SNFs had the greatest unadjusted change in mobility scores in overall patients (16.7 units and 16.6 units, respectively) and in musculoskeletal patients (19.4 and 20.7 units, respectively). HHA patients had unadjusted mobility change scores of 12.1 overall and 16.9 in musculoskeletal patients. Among the more complex nervous system disorder patients, those treated in IRFs achieved 14.8 units improvement, those treated in SNFs achieved 12.6 units improvement, and LTCH patients improved 11.2 units, followed by HHA patients with 10.4 units change. These results are not adjusted for variation in patient characteristics.

Differences in mobility at discharge were examined using multivariate models that controlled for patient acuity characteristics at admission. In these models, after controlling for differences in populations admitted, provider setting did not have a significant effect. This suggests that the differences seen in the unadjusted rates can be accounted for by patient characteristics and the severity of the populations admitted to each setting. This finding was also seen in the condition-specific models tested. In summary, key findings related to the prediction of change in mobility functioning are as follows:

- After controlling for patient acuity, the provider setting is not a significant predictor of change in mobility from admission to discharge.
- The nonsignificance of setting in predicting change in mobility held when the two diagnosis subpopulations of interest were examined.

ES.6.2.3 Hospital Readmission within 30 Days of Discharge

The third outcome examined was 30-day hospital readmissions. This was a key outcome for considering the impact of medical treatments on returning the patient to a better health status. Within the sample, unadjusted readmission rates within 30 days of hospital discharge were similar across provider types. The overall rate of readmission in the sample was 19.2 percent. IRFs had the lowest proportion of patients in the sample who were readmitted (17.4 percent), followed by SNFs (19.8 percent), HHAs (20.2 percent), and LTCHs (21.1 percent).

After adjusting for patient acuity at the time of admission to the PAC setting, patients in LTCHs appear to have lower probabilities of readmissions within 30 days of discharge from the initial acute hospital relative to SNFs. No significant differences were found between IRF or HHAs and SNFs in the adjusted probability of 30-day hospital readmissions. It is important to note that this analysis did not attempt to examine the cause of readmission or the patient acuity level at the time of readmission. The four PAC settings vary in their capacity to treat emergent medical situations, and the level of acuity that may trigger a readmission will be different in an organization that is classified as an acute hospital (including LTCHs) compared with a subacute provider (including SNFs). Thus, the lower readmission rate found in LTCHs is an anticipated reflection of their status as a hospital.[4]

Key finding for readmission analysis:

- After controlling for patient acuity differences at admission to the PAC setting, LTCH patients appear to have significantly lower probabilities of being readmitted to the acute hospital within 30 days of discharge relative to an SNF setting. The capacity of LTCHs to deal with higher severity patients may be associated with this finding.

[4] Subsequent analysis found that while readmission rates were lower for LTCHs in the 30 days since acute discharge, rates in days 31-60 were higher than for cases treated in other PAC settings (ASPE, 2011).

ES.7 Resource Use—General Methods

Sections 9-11 in Volume 4 of the report present the information related to the analytic approach and results associated with the resource intensity index analysis. Section 9 presents the information related to the analytic approach and extensive descriptive information about the sample used in the resource intensity analysis. It briefly reviews preliminary resource models results presented in the RTC associated with this project. Regression methods were used in this analysis. Section 10 describes further exploration of the model structure and new variable formulations using a different statistical approach, classification, and regression trees. Section 11 presents the results of integrating the findings of Section 10, new variables, and their formulations, with the regression methods, which are more usable for final modeling.

ES.7.1 Analytic Approach

CARE assessment data and CRU data were used along with data from claims to perform analyses predicting resource use in the four PAC settings. The basic measure of resource use is the weighted sum of total staff time per individual patient. Total staff time includes all direct care staff and support staff directly involved in the care of specific patients. Data were weighted to reflect each staff member's national wage rate by occupation and licensure level.

Two resource intensity index (RII) measures were constructed: one reflecting intensity of care provided by *routine,* nontherapy staff, such as nurses and aides (routine RII), and a second reflecting intensity of care provided by *therapy* staff, including physical, occupational, and speech pathology, to construct a therapy resource intensity index (therapy RII). These variables were modeled at the stay or episode level so that total resources for each patient were being modeled. This formulation puts a short stay with high daily intensity in one setting comparable to a stay in a setting with lower intensity but greater length of stay. In a home health episode a visit measure was used rather than days.

Unadjusted descriptive statistics were computed to profile the populations in each setting. The main analyses were done using regression approaches in which variables were constructed from the CARE and claims data to describe aspects of each patient's condition and explain the resource use measure. Resources were measured as the amount per stay or first HHA 60-day episode, and the amount per day.

One focus of the analysis was to determine which types of characteristics would be useful to explain variations in patient costliness or resource intensity. The other main purpose was to determine to what extent a consistent model could be used to predict resources across all settings, and, if complete consistency was not possible, what degree of inconsistency would be needed.

In examining the resource intensity models, several issues should be kept in mind. First, the data were collected using a sample framework designed to oversample certain key patient and provider characteristics. Therefore, the rates of patient use reported are not, and were not designed to be, reflective of the actual national population of patients treated in these settings. Second, the resource use information collected reflects the care that was provided within participating providers and does not necessarily reflect either ideal care or maximally efficient care.

The regression analyses modeled resource use with variables from the claims and CARE tool to determine which classes of variables seemed to be both statistically significant and substantive. But more important at this stage was to determine the degree to which the number of models could be collapsed while still achieving a reasonable fit. The relative performance of models was examined. Models were created with the following characteristics. A subset of these models was examined in the RTC materials.

- **All-PAC Settings.** This type of model estimates a single set of case-mix weights and a single base resource intensity amount for all PAC settings (HHA, IRF, LTCH, and SNF). This model predicts the intensity and amount of care for a given patient forcing the effects of the patient characteristics on intensity to be uniform across all settings.

- **HHA–Inpatient PAC Settings.** This pair of models is the same as the previous model, but it separates HHAs from inpatient PAC settings on the observation that home health resource intensity structures are significantly different based on the fewer hours of services being provided in the home setting.

- **HHA–LTCH–SNF/IRF.** This set of models allows the effects of patient characteristics on intensity in the HHA and LTCH settings to be unique to each of these individual settings. The effects of patient characteristics on intensity in the SNF and IRF settings are not allowed to differ from one another. This form of the model was not estimated in the original regression work.

- **HHA–Inpatient Diagnostic Groups.** This set of models allows the effects of patient characteristics on intensity in the HHA setting to be different from the effects of patient characteristics in the remaining settings. In addition, for the patients admitted to IRFs, LTCHs, and SNFs, it allows the effects of patient characteristics on intensity to vary across the following four broad diagnostic groups: neurological, orthopedic, respiratory, and medical/surgical conditions not otherwise categorized. This form of the model was not estimated in the previous regression work.

- **Setting-Specific.** This set of models allows each PAC setting to have its own set of case-mix weights and base resource intensity amount. The Setting-Specific models use consistent measures of patient acuity for each of the different settings, but this model is different from the other two models in that it allows the significance and impact of each measure to differ by setting.

ES.7.2 Sample Description Results

ES.7.2.1 Routine Intensity Descriptive Results

We found that the unadjusted, average routine resource intensity differed by setting in expected ways: the LTCH sample examined had the highest routine RII per stay, with about three times the staff resources per patient than in the IRF or SNF settings (161.4 RN-equivalent hours, compared with 58.6 and 50.9 RN-equivalent hours, respectively). HHAs had the lowest average nursing resource intensity per patient, with a mean routine RII of 5.3 RN-equivalent hours per 60-day home health episode). The lower numbers in HHAs reflect the nature of services in this setting where care is provided through visits rather than on a 24-hour basis as in an inpatient setting.

ES.7.2.2 Therapy Intensity Descriptive Results

Average therapy intensity per inpatient stay differed by setting. The stay-level unadjusted therapy intensity was greatest in IRFs, with a mean of 47.6 licensed therapist-equivalent hours per person per stay followed by a slightly lower stay-total in SNFs, with a mean of 43.9 therapist-equivalent hours per stay, and followed by LTCHs with 33.1 therapist-equivalent hours per patient stay.

The frequency of therapy care also varies across settings. On average, IRF patients received therapy on 5.2 days per week (or 74 percent of days), while SNF patients received therapy care on 4.3 days per week (or 62 percent of days). Therapy was provided to LTCH patients on 3.8 days per week (or 55 percent of days) on average. Roughly 52 percent of HHA days included some therapy.

ES.7.3 Review of RTC Regression Modeling Results

As an aid to the reader, Section 9 contains a brief review of the routine resource intensity models presented in the supplemental materials to the RTC discussion. This discussion acts as an introduction to the exploratory and model refinement work presented in Sections 10 and 11.

ES.7.3.1 Review of Routine Intensity RTC Regression Modeling

In the RTC models presented, patient acuity factors explained 63.6 percent of the variation in routine resource intensity across all settings in the All-PAC Settings model. In the HHA–Inpatient PAC Settings model, when HHA was separated from the three inpatient PAC settings, patient acuity factors explained 70.4 percent of the variation. Adding setting-specific indicators in the inpatient PAC component of the model only increased the explanatory power to 71.0 percent. (This overall explanatory power does not reflect the explanatory power of a model for each of the included settings viewed separately.)

The significance associated with setting-specific indicators was useful for understanding whether one or more payment models were needed if uniform acuity factors were used. The models that included setting-specific indicators suggested that HHA was significantly different from the inpatient PAC settings but that setting was not a significant predictor of routine resource intensity among the three inpatient PAC settings (LTCH, IRF, and SNF) after controlling for patient complexity. This suggests that HHA payment systems may need to be based on a significantly lower base rate than other settings, but the three inpatient PAC settings could use a common case-mix adjustment system.

Using the Setting-Specific model only improved the overall explanatory power slightly over the HHA–Inpatient PAC approach (mean square error [MSE] R-square of 73.5 rather than 71 as found in the HHA–Inpatient PAC model). While the use of four separate models, one for each PAC setting, could increase the explanatory power somewhat, the difference may not be enough to offset the advantages of having a system with greater cross-setting consistency in the case-mix model. Using the Setting-Specific model would result in each factor having a different impact across the four models; in other words, the coefficients would be reflecting setting-specific factors beyond those associated with the individual item. For example, the effect of a stage 4 pressure ulcer would be allowed to differ by setting, for reasons other than patient acuity factors.

The desirability of the HHA–Inpatient PAC Settings approach was further supported by the relatively low levels of under- or overestimation of these models. The average predicted routine resource intensity was within 10 percent or less of the actual intensity in each inpatient PAC setting, suggesting relatively little bias in the HHA–Inpatient PAC Settings models and further supporting the potential for moving toward one model for the case-mix adjustment component of the inpatient PAC payment systems. Further, this model explained much less variation in the HHA setting than in the inpatient settings, suggesting the possible need for more work to improve the HHA model.

In summary, key RTC findings related to the prediction of routine resource intensity included the following:

- Strong predictive models of routine resource intensity for the inpatient settings based on uniform definitions and measures of patient medical complexity across settings were created. This was accomplished with a limited set of patient acuity items defined in a common manner across each setting.

- Evidence supported the possible future development of a common case-mix adjustment system for the three inpatient PAC settings. This system would calculate the patient-specific resource expenditures portion of payment in the same manner across settings. These models can be created for all the three inpatient PAC settings with minimal over- or underprediction compared with actual resources use.

- Due in part to the nature of home health service provision of care, a payment model combining home health with the other types of PAC providers is not supported by the analysis. Many of the factors predicting routine resource intensity in HHAs were similar to the types of measures that were predictive of resource use in the other PAC settings. However, using one model in all four settings, with identical weights and base rates, would significantly overcompensate HHAs.

- Patient acuity measures that were predictive of routine resource intensity came from all three domains of the CARE Case Mix Classification Schema. This indicates that PAC payment systems can be improved by the inclusion of additional patient acuity measures found in the CARE tool, such as the addition of non-ICD-9 derived measures in LTCHs.

ES.7.3.2 Review of Therapy Intensity RTC Regression Modeling

The therapy resource intensity models had similar results to those seen in the routine models. Again, the HHA setting was significantly different from the three inpatient PAC settings. Separating HHAs from the inpatient PAC settings dramatically improved the explanatory power of the models without the additional need for setting indicators.

The All-PAC Settings therapy models had an overall MSE-based R-squared value of 0.249 when all settings were forced to have the same base rates and coefficients associated with patient acuity factors. The explanatory power increased to 0.343 for HHA and 0.360 for inpatient PAC settings when the two models were run separately in the HHA–Inpatient PAC Settings model. Adding setting indicators to the HHA–Inpatient PAC Settings models only increased the R-square by 0.017, suggesting that separate base therapy resource intensity amounts for each inpatient setting would only improve the model's overall explanatory power slightly. Therefore, as with the routine intensity, separating HHAs from the three inpatient settings was identified as a model with potential for further development.

Examination of the ratio of the predicted-to-actual therapy resource intensity shows that when HHAs are separated from the inpatient PAC settings, the potential for under- and overpayments varies by setting. Using the HHA–Inpatient PAC Setting model, the predicted therapy intensity for IRFs is within 1 percent of the actual intensity, SNFs are predicted low by 11 percent, and LTCHs are predicted 15 percent more than the actual value; LTCHs would be disproportionately overpaid using this model specification.

These findings suggested that it may be possible with a refined model specification to construct a payment model that pays providers fairly across settings by separating HHAs from the inpatient PAC settings while using a common set of case-mix weights and base resource intensity amount for the inpatient PAC settings. However, relative to the case for the routine resource intensity models, the challenges may be greater for the therapy intensity models since the across-setting bias is higher for LTCHs in the therapy RII models than in the routine RII models.

The results also support the use of separate nursing and therapy indices because the explanatory power of the routine and therapy models differed, although substantial levels of variation were explained in both. Treating nursing and therapy independently in the case-mix system will allow different factors to be used to explain variation in intensity and may improve the therapy intensity models.

In summary, key findings presented in the RTC related to the prediction of therapy resource intensity are as follows:

- Consistent payment models predicting patient-specific use of therapy services can be created for SNFs and IRFs with minimal bias. With additional work, these models might be revised to create consistent therapy use models that include all three PAC inpatient settings. Model results support modeling HHA therapy intensity separately.

- PAC payment systems can be improved by examining and modeling the therapy and routine patient-specific resource use separately.

- Good predictive models of therapy resource intensity based on uniform definitions and measures of patient functional complexity between different settings were created without the need for using measures of resource utilization.

The findings in this section suggested further exploration of models to understand better the use and formulation of the explanatory variables and alternate ways to stratify the population than just by setting. These findings are described in the next two sections.

ES.8 Resource Use—CART Analysis

Section 10 in Volume 4 uses a different analytical approach to study the determinants of resource use. The regression approach uses all explanatory factors simultaneously. In contrast, in the approach used in this section the explanatory variables are examined sequentially. We also investigate models that stratify the population by clinical characteristics rather than by setting.

ES.8.1 Analytic Approach

The approach described in this section is classification and regression tree (CART) analysis. In this technique the CART program is used to split the sample of interest into two subsamples based on values of the explanatory variable that best creates subsamples that are similar in resource intensity within each subsample and different between subsamples. Each of these subsamples is then split again, using values of the variable in the explanatory variable set that produces the best split of each subgroup, usually a different explanatory variable. Each split is conditional on the splits made previously. The average value of the resource intensity for each group is the prediction of the resource intensity index for the members of that group. The variables that have the most power in creating splits are of most interest.

The sequential splitting allows an examination of the variables that seem to have the most power after other variables have been used in the splits. For numeric variables, as opposed to variables that are categorically yes or no, we can observe the values that are used to divide the sample. The models can also be evaluated for explanatory power and biased prediction.

ES.8.2 Results

Many models were explored using the regression tree approach including the formulations in the prior section and models that were not built on setting but on clinical characteristics of the beneficiaries.

Specifications 1 and 2. In these specifications, observations from all four settings were pooled. Specification 1 included setting indicators and Specification 2 did not.

Specifications 3 and 4. In these specifications, observations from the three inpatient PAC settings were pooled. Specification 3 included setting indicators and Specification 4 did not.

Specifications 5 to 8. In these specifications, observations from each PAC setting were examined separately. The specification samples were, in order, HHA, LTCH, SNF, and IRF.

Specifications 1 through 4 are most useful when considering the ability to create a successful single model that can explain variation of each RII within multiple settings. Specifications 5 through 8 are useful in considering the question of whether the same factors are important predictors of the RIIs across settings.

Four additional specifications were also considered. In these specifications, the inpatient PAC observations were stratified into four broad groups based on primary diagnosis. HHA cases were not modeled under these specifications. The objective was to create broad primary diagnostic groups based on the primary medical, surgical, or injury-related diagnoses for which patients were originally hospitalized. The grouping strategy for these diagnoses was to combine conditions expected to cause similar disabling impairments. Consequently, each group has diagnoses that affect the function or structure of similar organs, thus having similar effects on how they regulate the ways and manners in which people can perform self-care, on mobility, and on cognitive activities. Diagnostic groupings have the characteristic of not being setting specific but do have some correlation with the various settings that treat such patients. The four broad diagnostic groups were the following:

Specification 9: Inpatient PAC, Neurologic Conditions. This group includes patients with one of the three neurological primary diagnoses: stroke, along with medical and surgical neurologic diagnoses.

Specification 10: Inpatient PAC, Orthopedic Conditions. This group includes patients with a primary diagnosis in one of the five orthopedic categories: minor and major orthopedic surgery, minor and major orthopedic medical diagnoses, and conditions related to the spinal column.

Specification 11: Inpatient PAC, Respiratory Conditions. This group includes patients with primary diagnosis in one of the four respiratory categories: ventilator/tracheostomy, COPD, respiratory surgeries, and other medical diagnoses related to the respiratory system.

Specification 12: Inpatient PAC, Other Medical/Surgical Conditions. This group includes patients with a primary diagnosis that does not belong to any of the other three categories.

As part of the exploratory process a number of refinements to the patient acuity measures were examined. One example is a case complexity summary measure created from the set of comorbid conditions in the risk adjuster. This is a numeric score depending on how many and which comorbidities are present.

We report here the salient findings from these models. The term “importance” refers to the relative explanatory power of a variable in explaining the RII. Substitutability refers to the degree to which a variable can serve as a substitute for the variable actually used to make a split.

ES.8.2.1 Routine Intensity

The relative importance of the specific patient acuity measures in each of the four setting-specific models (Specifications 5-8) of the routine RII was examined to compare the important variables across settings. The most striking result is that the top three most important predictors of routine RII for LTCH stays have no relevance in the IRF and SNF settings. Length of ICU stay is by far the most important factor in explaining variation in routine RII in the LTCHs, followed by ventilator treatment and the primary diagnosis of ventilator/tracheostomy. Given that there are so few patients in the SNF and IRF sample with an ICU stay, this variable was not found to be an important splitter in the CART analyses for these settings.

The Rasch mobility and self-care scores and the comorbidity index are among the variables that are important predictors of routine RII across all settings.

The relative importance of the specific patient acuity measures in predicting routine RII for inpatient PAC stays within each of the four diagnostic groups were also examined (Specifications 9-12). For the neurologic and orthopedic patients, the Rasch function scores and comorbidity index play the three biggest roles in driving the routine RII. The self-care score is most important for patients with the neurological conditions, where one would expect upper and lower extremity involvement. In contrast, for orthopedic patients where lower extremity impairment tends to be predominant—due to large numbers of people with joint replacement and hip fracture—the mobility score is slightly more important than the self-care score. All three variables also help explain variation in the routine RII for the other diagnosis groups.

Length of ICU stay is by far the most important factor in explaining the routine RII for the other medical/surgical and respiratory patients. The second most important factor in explaining variation in the routine RII for other medical/surgical patients is the comorbidity index. For respiratory patients the second most important variable is "no intake by mouth," with having an indicator of ventilator use only slightly less important.

In comparing the setting-specific models to the clinical models there are some parallels and differences. The importance of ICU stay length and ventilator status comes up in the respiratory model and the LTCH model. The populations associated with severe respiratory patients, long stay ICU patients, and LTCH patients tend to overlap. Although there are medical/surgical patients in the LTCH and IRFs, the ICU stay is not very important in those settings despite being an important factor for that clinical group. The longer ICU patients tend not to be found in those settings. The advantage of the clinical grouping is that the predictors are based more on the basic patient characteristics across settings than the setting alone.

The patterns of the splits in the modeling were also suggestive of other potential alternative forms of the models. In models with setting variables, the HHA group is split from the others at an early split. However, LTCH also splits off at an early stage, with the SNF and IRF split not occurring till far down the tree of splits. This suggests an SNF/IRF model as an intermediate stage between all inpatient and setting specific.

In comparing an all inpatient model with setting indicators to one with clinical variables only (Specifications 3 and 4), CART uses other acuity variables (e.g., ICU days and ventilator use) as substitutes for the LTCH setting to achieve almost the same explanatory power. This does support the possibility of an all inpatient model. The R-squared statistics for these models are comparable to those in the regression analyses in the prior section.

ES.8.2.2 Therapy Intensity

The relative importance of the specific patient acuity measures in each of the four setting-specific models (Specifications 5-8) was examined to compare the important variables across settings for therapy. The Rasch self-care and mobility scores are generally very important drivers of the therapy RII for all settings. This pattern of importance is expected, though the relative importance in predicting amount of therapy varies between and within settings. In LTCHs the marker for being a ventilator patient is important; it is not important in the other settings. Age and the comorbidity index are also important markers of therapy intensity across the settings. Some factors differ in importance across settings. Stroke is important in the HHA and IRF settings only. Sitting endurance is important only in the LTCH and IRF settings.

In the models for therapy RII in inpatient PAC stays within each of the four diagnostic groups (Specifications 9-12), the mobility and self-care scores are important across all conditions. In addition the comorbidity index and age are generally important. These are similar in pattern to the setting -specific models. A past stroke is important in predicting therapy RII in the respiratory, orthopedic, and neurologic groups. Sitting endurance is important in all but the respiratory group. Some of the factors that are very important in the routine care models are not found to be determinants of therapy (e.g., ICU days and ventilator use).

The therapy models have poorer explanatory power than the routine care models, which is consistent with results found in prior work. An exception is the HHA model, which has slightly better explanatory power for therapy RII. The modeling, as usual, indicates that the HHA should be split off. For the models of all inpatient settings (Specifications 3 and 4), the addition of setting indicators improves the explanatory power by a few percentage points measured by the R-squared statistic. Again there is some evidence that SNF and IRF could be in one model separate from the LTCH.

ES.8.2.3 Suggested New Explanatory Variables

The CART software indicates variables that are close substitutes for one another in differentiating groups of patients. One of the indications is that the self-care and mobility function scores can act as substitutes for one another. In response, we created a combined motor score to use in the next set of analyses. Another group of substitutes is the primary diagnosis of tracheostomy with ventilator therapy in the prior acute stay, with the ventilator use in the PAC setting. The primary diagnosis of ventilator/tracheostomy has been subsumed into the respiratory surgical group. The ICU stays and the comorbidity index are good substitutes for the LTCH setting marker.

Other changes made for the next regression analyses suggested by the splits at various values of the continuous variables was to allow those variables to have different incremental effects depending on whether they have high or low values. The ICU days and comorbidity index would be tested in the linear and squared form, so the effect of a unit change could be different at low levels and high levels.

ES.8.3 CART Analysis Conclusion

The process of studying the relationships of the variables with a different analytical method was useful in going back to the regression formulation with some new variations in the modeling. The use of the actual CART models created would be problematic with the sample sizes available. The process of splitting and resplitting patient groups leads to small samples in the final nodes. The precision and external validity of models built on small samples is not adequate for final modeling, though it is suggestive in the model building. The next section of the report returns to regression models in light of the insight obtained from the RTC findings and these exploratory results.

ES.9 Resource Use—New Regression Analyses

With the information from the early regressions and the suggestions from the CART analysis, we approached the regressions in Section 11 with some new variables, including the comorbidity index and continuous and nonlinear variables. Two additional models were examined in addition to the three explored earlier: The HHA–LTCH–SNF/IRF model and the HHA–Inpatient PAC Diagnostic Group model. The results of the regressions are reported in great detail in Section 11. We will concentrate here on the main findings of the reformulation of the models and the new variables used.

ES.9.1 Routine Resource Intensity Index

ES.9.1.1 The All–PAC Model

The All–PAC model, as before, was not satisfactory because including the HHA with the inpatient settings yields the lowest global R-squared among the five models being compared. The All–PAC Settings model also provides fairly biased predictions for the routine RII across all settings; the bias is most pronounced for HHAs. Here the model, on average, predicts a routine RII that is 3 times greater than the actual value in this setting. At the same time, the model underpredicts the routine RII in the three inpatient PAC settings. It underpredicts routine RII by more than 25 percent in SNFs and by roughly 17 percent in IRFs and LTCHs.

ES.9.1.2 HHA–Inpatient PAC Settings Model

In the HHA–Inpatient PAC Settings model the results improve significantly. For example, the R-squared increases from a negative number[5] to 0.141 for HHA episodes, and it improves from 0.033 to 0.093 for SNF stays. The bias in the three inpatient settings is never greater than 10 percent. For instance, the model overpredicts the routine RII by 9.4 percent for IRF stays and by 7.7 percent for SNF stays. The model underpredicts the routine RII in LTCHs by roughly 8 percent. Adding setting indicators adds little to the explanatory power, though it, as always, adjusts the mean predictions for each setting so there is no average bias.

Concentrating on the new variables used in the model we found for the inpatient component, longer ICU stays are associated with a higher routine RII in the inpatient PAC settings model, although the impact of this variable diminishes as ICU stays get longer. The inclusion of the squared term for ICU stay allows us to see this effect. Length of ICU stay is not significant in the HHA components of the model. The comorbidity index is significant only in the HHA intensity component of the model where a higher index is associated with a higher routine RII. The squared term indicates that this relationship diminishes somewhat as the index increases. Higher functional status, as measured by the Rasch motor function scale, is associated with a lower probability of receiving routine services in the HHAs. However, among patients who received any services, the relationship between functionality and the routine RII is positive for patients with a Rasch score below 35 and negative for patients with higher scores. The result is different for the inpatient PAC settings, where the relationship between functional

[5] Computation of the MSE-based-R-squared for a subpopulation in a model built on a larger population can result in negative numbers if the fit is poor for that subgroup.

performance and the routine RII is negative for all patients with a Rasch motor function core of greater than zero.

ES.9.1.3 HHA–LTCH–SNF/IRF Model

The HHA–LTCH–SNF/IRF model is a new model set that was developed based on results from the CART analyses. It was suggested by results that indicated that the IRF and SNFs were more similar than they were to the LTCHs in modeling the routine RII. By removing the inherent constraint in the all-inpatient model that the coefficients and base rates be the same for LTCH and for IRF/SNF, the fit for IRF and SNF stays improves dramatically. For IRFs the R-squared increases from 0.249 to 0.381. The R-squared for the SNF stays more than doubles from 0.093 to 0.223. With respect to the predictive ratios, because LTCH stays are being modeled separately, the predictions for the routine RII are unbiased for this setting. But the predicted-to-actual ratio also improves significantly for the IRFs, falling from 1.094 to 1.016. The predicted routine RII is also less biased for SNF stays, as the model predicts routine RII that, on average, is 2.5 percent less than the actual as opposed to 7.7 percent higher in the prior model. These overall improvements in predicting for settings are to be expected as this structure approaches the setting-specific model that optimizes the fit of the models for each setting.

Among the newer variables added, the comorbidity index in HHA behaves as in the HHA model above; it is not significant in any of the settings after controlling for the comorbid indicators and the other patient acuity measures included in the models. For the combined SNF/IRF settings, the relationship between motor function, as measured by the Rasch motor function scale, and routine intensity is positive at lower levels of function, but becomes negative at higher levels of function. Thus, for most of the patients, higher functional level is associated with a lower intensity of routine care. In the LTCH as well there is a diminishing effect with higher function. The ICU days has a positive relationship to the RII in the LTCH without an effect from the squared term.

ES.9.1.4 HHA–Inpatient PAC Diagnostic Groups Model

The HHA–Inpatient PAC Diagnostic Groups model is conceptually different from the others in that it does not divide the sample by setting, but by clinical groupings. Although the patterns of care in each setting are important, the model does not attempt to fit by setting but by patient characteristics. The patients are divided into four strata: neurological, orthopedic, respiratory, and not otherwise classified medical/surgical cases. The effects of the settings may come in if a setting treats more or fewer patients of a given type. It is of interest to determine how well the model predicts for each setting using clinical splits rather than modeling all patients in one equation or setting-specific equations. The results of the stratifications described are for the inpatient settings only. HHA patients were modeled separately.

Overall, the model is an improvement in performance over the HHA–Inpatient PAC Settings model. The R-squareds improve significantly for each inpatient PAC setting, from 0.249 to 0.316 for IRFs, from 0.619 to 0.699 for LTCHs, and from 0.093 to 0.180 for SNFs. Also, the predictions carry less bias than those in the HHA–Inpatient PAC Settings model. The predicted-to-actual ratio improves from 1.094 to 1.077 for IRFs, from 0.921 to 0.941 for LTCHs, and from 1.077 to 1.047 for SNFs.

Whether the model is statistically an improvement over the HHA–LTCH–SNF/IRF model is less clear. The global R-squared is better (0.788 as compared with 0.769), but the R-squareds are worse for the IRF and SNF stays. Additionally, the predictions of the routine RII are more biased for the IRF stays in the HHA–Inpatient Diagnostic Groups model; the predicted-to-actual ratio is 1.077 as compared with 1.016 in the HHA–LTCH–SNF/IRF model. The predictions are also more biased for the SNF stays. Finally, the improvement in the fit for LTCH stays is countered by the introduction of some bias because the LTCH does not have its own model or setting indicator.

The inclusion of setting indicators in the HHA–Inpatient Diagnostic Groups model increases the global R-squared by only 0.007 from 0.788 to 0.795. Thus, the setting factors explain very little beyond the case-mix factors, suggesting that separating HHAs from the inpatient PAC settings and, for the inpatient PAC settings, allowing the effects of patient characteristics on the routine RII to vary across the four broad diagnostic groups improves the explanatory power of an all-inpatient PAC model without the need for setting indicators.

As will be discussed later, the importance put on deriving the best fit for each setting, conditional on current practices, is only one factor in determining which multisetting model is a reasonable approach. On its own terms, the model estimated on clinical strata has R-squared values of 0.350 for the neurologic group, 0.415 for the orthopedic group, 0.714 for the respiratory group, and 0.662 for the medical/surgical group. This indicates that without any setting information per se, reasonable explanatory power is possible when stratifying by patient diagnosis type, with the neurological group more challenging than the others.

Length of ICU stay is important in the medical/surgical, orthopedic, and respiratory components, with greater length of stay associated with a higher routine RII. In the orthopedic component, the result on the squared term for length of ICU stay indicates that the impact of this positive effect on the routine RII diminishes for longer ICU stays.

Increased functional ability, as measured by the Rasch motor function scale, is associated with a lower routine RII in the not otherwise classified medical/surgical component at all levels of function. For the neurologic patients, increased function leads to a higher routine RII at low levels of function (where the Rasch score is less than 20) but to a lower routine RII at higher levels of function. For the orthopedic patients, increased function leads to greater routine intensity for patients with Rasch scores below 17 but leads to less routine intensity at higher level of function. Motor function at admission is not significant in the respiratory component.

The new comorbidity index had effects only in the orthopedic model. The squared term indicated that higher comorbid levels were associated with lower RII. The coefficient on the interaction between the Rasch score and comorbidity index in the orthopedic component suggests that at higher level of function the relationship between increased comorbidity and the routine RII is positive.

ES.9.1.5 Setting-Specific Model

The Setting-Specific model is customized to each setting and fits the patterns of care as currently delivered by allowing the base rate and the coefficients on every explanatory variable to be customized to that setting. As would be expected, such a model is not biased for any setting, because the method always has a predictive ratio of 1.0 when the predicted sample and the estimation sample are the same. With complete customization by setting, this model also improves the fit for IRFs and SNFs as compared with the HHA–Inpatient PAC and HHA–LTCH–SNF/IRF models. The improvement in the fit for SNF stays is substantial; the R-squared improves to 0.377 for this setting compared to 0.223 when it was combined with IRF. The improvement in the fit for IRF stays is less pronounced but still significant; the R-squared rises to 0.424 from 0.381 when combined with SNF.

As for the new variables formulations, the number of days in the ICU is significant only in the LTCH model; the squared variable is not significant so does not increase or decrease the incremental effect as ICU days increase. The comorbidity index is not significant when each setting is modeled separately. For the IRF and SNF settings the relationship between motor function, as measured by the Rasch motor function scale, and routine intensity is positive at lower levels of function, but becomes negative at higher levels of function. In the LTCH the effect of higher motor score is only to reduce intensity of routine care.

ES.9.1.6 Routine RII, Summary

In reviewing the models with a focus on the variables that have significance within each setting under current care patterns, one sees that among the uniform set of explanatory variables applied to each setting and model, there is variation as to which are of greatest importance across settings. In some cases the same variable has opposite directions in different settings, for example, a primary diagnosis of COPD or cardio-surgical has negative implications in IRFs and SNF and positive implications in LTCHs. This means that inpatient models that span settings, such as the HHA–Inpatient PAC Setting or the HHA–Inpatient PAC Diagnosis stratified model, have to estimate coefficients that are compromises to best fit across the settings. Some of the variables that have significance in the spanning models but not in the setting models, such as the comorbidity index, serve to capture some of the effects of setting but through an additional clinical measure. Another complicating factor interpreting individual coefficients is that some settings have low frequencies of the characteristics resulting in higher variances and lower statistical significance.

The bias shown when looking at predictive ratios for a particular setting within a spanning model represents resource intensity that the clinical aspects of the model cannot predict. The bias carries setting-specific base care patterns that differ by setting. The models that span the inpatient settings underpredict for the LTCH as the models have to capture the base resource intensity in the SNF and IRF settings. Putting setting indicators into the models can capture the existing differences if desired. The clinical explanatory power of the models to differentiate resource intensity by patient characteristics changes only slightly.

ES.9.2 Therapy Resource Intensity Index

The models estimated for therapy resource intensity are the same in form as those for routine resource intensity. The dependent variable was the therapy RII and the explanatory variables were the same set as used for routine services. In general, the R-squareds for all the therapy models were lower than those for models predicting routine care.

ES.9.2.1 The All–PAC Model

As with the routine RII models the therapy RII model with all settings included was the least powerful. The overall R-squared was 0.281. The MSE-based R-squared is negative for HHA episodes, indicating a very poor prediction. It is 0.043 for LTCH stays, 0.040 for SNF stays, and 0.158 for IRF stays. The biases also indicated problems with keeping all the settings in one model. The overprediction for HHAs was 37 percent.

ES.9.2.2 HHA–Inpatient PAC Settings Model

By separating the HHA setting the global R-squared increases from 0.281 to 0.356. The R-squared for three of the individual settings also improves. For example, it increases from a negative number to 0.179 for HHA episodes, from 0.158 to 0.186 for IRF stays, and from 0.040 to 0.129 for SNF stays. However, this model fit the LTCH observations more poorly, with an R-squared of 0.028 as compared to 0.043 in the All-PAC Settings model. The R-squareds are relatively week for therapy.

With respect to the new variable formulations, in the inpatient PAC component, longer ICU stays are associated with a lower therapy RII, although this negative effect diminishes as the length of ICU stay increases and becomes positive at roughly 4 weeks. The squared term of the ICU days produces this effect. The other new variables have significance as well. The fact that the squared term on the Rasch motor score is less than one indicates that increased functional status is associated with a lower therapy RII in the HHAs and that this relationship becomes more pronounced at higher levels of function. In the inpatient PAC settings it is related to a higher therapy RII at relatively low levels of functional ability. However, this relationship becomes negative at higher levels of functional ability. For instance, in the inpatient PAC settings, an increase in the Rasch motor score from 10 to 11 would increase the therapy RII by 1.3 percent, while an increase in the Rasch motor score from 20 to 21 would decrease the therapy RII by roughly 1 percent. At low functional levels higher scores are related to more therapy, while at higher levels of function a higher level of function reduces therapy.

The result on the interaction term between the comorbidity index and the Rasch motor score for the HHA patients indicates that the effect of increased comorbidity on the therapy RII is less pronounced, and even becomes negative at higher functional levels. As an example, for an HHA patient with a Rasch motor score of 20, a one-unit increase in the comorbidity index would be related to a 2 percent increase in therapy RII, while for an HHA patient with a Rasch motor score of 30, a one-unit increase in the comorbidity index (poorer health) would be related to a 2 percent decrease in therapy RII.

ES.9.2.3 HHA–LTCH–SNF/IRF Model

In the HHA–LTCH–SNF/IRF model the settings are further split. Only the SNF and IRF settings are constrained to have the same coefficients. The R-squared for IRF stays improves from 0.186 to 0.225, and it improves slightly from 0.129 to 0.132 for SNF stays. The R-squared improves dramatically for LTCH stays, which has its own model increasing from 0.028 to 0.237.

The predictive ratios for the RII are 1.0 for the settings having their own models. But the predicted-to-actual ratio actually is worse for the IRF stays than it was in the HHA–Inpatient PAC Setting model. Here the overprediction rises from a little less than 1 percent to 4.4 percent.

At the same time, the underprediction for SNF stays falls from roughly 9 percent to 6.6 percent. The bias for each individual setting depends on the relative proportion of patients in each setting that is included in the model.

The length of ICU stays has an impact only for LTCH patients where increased ICU stays are associated with a lower therapy RII. This impact diminishes slightly with the length of the ICU stay for LTCH patients. Length of ICU stay is not a significant predictor in the combined IRF and SNF model. The comorbidity index is not significant in the HHA, LTCH, or SNF/IRF models.

ES.9.2.4 HHA–Inpatient Diagnostic Groups Model

The HHA–Inpatient Diagnostic Groups model is the second model that combines all the inpatient PAC settings. It does stratify the model by primary diagnosis clusters allowing for somewhat more subtle distinctions among patients. Overall, the model is an improvement over the HHA–Inpatient PAC Settings model. The R-squareds improve significantly in each inpatient PAC setting , and the global R-squared improves from 0.356 to 0.460. Also, the predictions carry somewhat less bias than those in the HHA–Inpatient PAC Settings model. The predicted-to-actual ratio improves from 1.008 to 1.007 for IRFs, from 1.118 to 1.091 for LTCHs, and from 0.908 to 0.928 for SNFs. The stratification does not capture all the setting-specific patterns that are captured in setting indicators. Adding setting indicators in this therapy model does improve the R-squareds, increasing the global R-squared by 0.017 from 0.460 to 0.477. The setting patterns are more powerful in determining therapy than routine care.

Whether the model is an improvement over the HHA–LTCH–SNF/IRF model is less clear. The global R-squared is better (0.460 as compared with 0.387), as are the R-squareds for IRFs and SNFs, but the R-squared is worse for the LTCH stays, falling from 0.237 to 0.130. Additionally, the predictions of therapy RII are biased for the LTCH stays in the HHA–Inpatient PAC Settings Diagnostic Groups model (the bias is 9 percent), while being unbiased in the HHA–LTCH–SNF/IRF model, which models LTCH on its own.

On its own terms, the therapy model estimated on clinical strata has R-squared values of 0.299 for the neurologic group, 0.347 for the orthopedic group, 0.307 for the respiratory group, and 0.174 for the medical/surgical group. For therapy, the medical/surgical group is the most challenging to model. While these are lower than the comparable model for routine RII, all the therapy models have lower R-squares.

For the new variables, length of ICU stay previous to PAC admission has a significant effect in the medical/surgical and respiratory components. The results on the squared term indicate that the marginal effect is generally negative for ICU stays of less than 1 month but that the effect becomes positive for cases where very long ICU stays were involved. These components and long ICU stays are associated with LTCH stays.

As is the case for routine RII, the comorbidity index is significant. In the medical/surgical, orthopedic, and respiratory models, a higher index is associated with less therapy intensity. At the same time, the relationship between the index and the therapy RII varies across the models and the relationship changes at different points in the index. For the average medical/surgical patient (the index is equal to 2.3, on average, for these patients), a

1-unit increase in the index would be related to a 7 percent decrease in the therapy RII. For the average orthopedic patient, who has lower complexity (the index is equal to 1.5, on average, for these patients), a 1-unit increase in the index would be related to a 9 percent increase in the therapy RII.

Functional ability as measured by the Rasch motor scale is significant in all of the models. In each model, the marginal effect of increased functional ability as measured by the Rasch motor function score is positive at lower levels of functional ability (where the Rasch score is less than 30), but negative at higher levels of ability (roughly where the Rasch score is greater than 30). The significant coefficient less than 1.0 on the squared motor score coefficient creates the negative effect.

ES.9.2.5 Setting-Specific Model

The Setting-Specific model is the optimal model, if producing the best fit by setting is the main criterion. The predictive ratio for each setting is 1.0 by construction, and the coefficients are customized to each setting's current resource intensity patterns.

The Setting-Specific model has the best fit of all the models considered, with an R-squared of 0.463. The improvement in the fit for SNF stays is substantial; the R-squared improves to 0.306 for this setting. The improvement in the fit for IRF stays is less pronounced but still significant; the R-squared rises to 0.302. When compared to the HHA–Diagnostic Groups model, the overall fit is roughly the same (the global R-squareds are 0.463 and 0.460). Also, the fit for IRFs is the same (the R-squareds are 0.301 and 0.302). The fit for SNFs is slightly better in the HHA–Inpatient PAC Diagnostic Groups model, but the fit for LTCHs is much better in the Setting-Specific model (an R-squared of 0.237 as compared to 0.130). It is interesting that this therapy model is not notably better than the diagnosis stratified model except for LTCHs.

The length of ICU stays has an impact only for LTCH and IRF patients where increased ICU stays are associated with a lower therapy RII. This impact diminishes slightly with the length of the ICU stay for LTCH patients. The comorbidity index is related to the therapy RII only in the SNFs. For SNF patients, a higher comorbidity index is related to a lower therapy RII.

The impact of functional status in LTCHs, as measured by the Rasch motor function score, becomes increasingly associated with reduction in the therapy RII as functional status improves. For SNF and IRF patients, increased functional status is associated with a higher therapy RII at relatively low levels of functional ability. However, this relationship becomes negative at higher levels of functional ability.

ES.9.2.6 Therapy RII, Summary

Much that was said in the context of the routine RII models is also true for therapy. While the explanatory power is lower for therapy, in the context of fit for individual settings, the clinically stratified model is the best model that spans multiple settings. Better fits for each setting require isolating particular settings. The somewhat weak explanatory power found for therapy RII might be explained, at least in part, by the regulations governing each setting and the varying incentives in payment systems. IRF patients should receive 15 hours of therapy a week on average and, thus, the impact of clinical factors in determining therapy intensity will be reduced because the majority of patients may receive a similar level of treatment. Payment systems also currently include increases in payment because of explicit or implicit measures of therapy given the patients. Home health payments count therapy visits in determining payment. These factors make purely clinically based models more challenging.

ES.9.3 Weighted Regressions

Section 11 of the report also describes results of the same set of models with the observations in the models weighted by the proportion of patients in the settings. In the analysis sample used in all the analyses reported thus far, there was considerable oversampling of patients in LTCHs and IRFs. This was needed to better understand these populations. A simple random sample would have had just a few percent of such patients. The estimates generated in the models are driven to a degree by the proportion of HHA, SNF, IRF, and LTCH patients in the model. The coefficients reflect the character of the sample. We then reanalyzed the data with weights representing the proportion of SNF, IRF, and LTCH stays in the national claims data.

As would be expected, the very heavy reweighting of the SNF stays resulted in better explanatory power in the SNF setting and poorer fits in the IRF and LTCH settings. The essential finding from the weighted analysis is that the models can be customized by weighting the settings. It is possible to improve the explanatory power for a setting with a relatively low R-squared by sacrificing some of the power in another setting. This allows the models to have a more even predictive power by setting without using explicit setting information. Prospective models used for payment rarely predict exactly correctly for any individual case. The ultimate criterion is how they predict on average for cases of a given type. The better the explanatory power, the more likely it is to get good average payment with smaller numbers of cases. Weighting is one tool to even out the predictive power.

ES.10 Overall Conclusion for Resource Intensity

The modeling done in this work has shown that a uniform set of data sources, claims, and the CARE assessment instrument can be used across all the settings. It has also been demonstrated that it is reasonable to create models covering multiple settings. We find that segregating the HHA setting, which can be considered outpatient, provides better results for both the HHA and the inpatient PAC settings that remain.

Our evaluation of the models is contingent on the underlying data that reflect the practices driven by current conditions of participation of the provider types and payment systems. The implication is that choosing the models with the best explanatory power is allowing the past to drive the future. In a system in which similar types of patients may be served by different types of providers or multiple types of providers, it would be preferred to create payment systems with a degree of commonality. Settings do differ by the range of services they are expected to provide under current law. SNFs do not cover the same range of ancillary services as hospitals, for example. Differences in facility overhead and required staffing are also present. A final payment model would have to recognize the differences that remain in the systems.

We expect the models pursued to be able to be updated without major restructuring. For example, possible changes in the rules for LTCH patients in the mean length-of-stay requirement or for IRFs in therapy requirements could change the mix of patients. Models such as these that are strongly based on patient characteristics may need relatively little adjustment to the extent the models cover patients across settings. Setting-specific models might need more adjustment.

If a single approach is used that models the inpatients settings as a whole, the all-inpatient model works best when stratified into clinically coherent groupings. It is true that stratification increases the total number of parameters to be estimated and therefore the likelihood of better explanatory power. However, the stratification does not simply add more parameters; it is a logical approach to improving the model. By putting patients into classes of patients with similar conditions, there is a greater likelihood that the risk adjusters in each model will work better within each class than they would if constrained to be the same across all classes.

If any of these approaches are used, the nature of the development would be to (1) refine the clinical covariates in the models and (2) if a medical condition stratification is used, refine the definitions and extend the modeling to ancillary services and methods to combine the separate components of the prediction models.

ES.11 Conclusions

The PAC-PRD was very successful in developing collaboration among the many types of providers treating the Medicare population and gaining input from the many clinicians working with these cases on a daily basis. One of the greatest contributions of this effort may be the consensus-based development of standardized items for measuring medical, functional, and cognitive complexity regardless of site of care. The CARE items were tested with both the acute and PAC populations and showed high reliability in each setting. This finding is not surprising, because these items are each already being used in one or more settings by clinicians who frequently work in more than one setting during their career.

Individual clinicians gave positive feedback on the items. They appreciated the input of national experts in developing a standard way to measure pressure ulcers and other factors. Those working in SNFs and HHAs appreciated the greater specificity of the function measures tested in this demonstration. The measures built on the existing science, but with the input of clinical rehabilitation experts, modified the underlying Barthel scale to allow measurement of populations with a wider range of function, from the most impaired in the LTCH setting, to the least impaired in some of the HHAs.

The results of the analyses show that consistent measurement across settings is possible. The resource intensity sections showed that the case-mix items in the CARE item set explain variation in resource use, and very little was added by inclusion of the indicator variable for the setting after separating out home health. This is a very important finding in that it shows that patient characteristics can be used to explain costliness. More importantly, having consistent ways of measuring these characteristics allows comparison of outcomes to understand the relative impact of treatments provided in different settings.

The results also underscored that these settings and the populations treated in them are not mutually exclusive. Similar patients may be treated in different types of settings across different market areas. Having a consistent assessment approach will allow CMS to begin considering the best approaches for refining the PAC payment systems.

The CARE tool was designed as a set of items that could uniformly measure concepts already largely included in the different PAC PPSs. The implementation of CARE within the demonstration was successful. All five settings were able to use the CARE items to collect information in a consistent, reliable, and comprehensive manner for their Medicare populations. Participant feedback on CARE was generally positive, with support from each clinical community for CMS' effort to use nationally accepted standards, as in the case of the pressure ulcer development, or to improve on weaknesses in the current measures, as in the functional status items. The CARE function items addressed some of the ceiling and floor effects associated with the current assessment instruments and provided greater specificity for measuring change than the current MDS and OASIS function items.

Reliability testing for the CARE items showed that these items met the same standards of reliability as the current CMS-mandated patient assessment items. Overall, the inter-rater reliability results showed very good agreement on most items, suggesting that these items could be used to measure a patient's progress in a standardized way across an episode of care.

The development and testing of the CARE tool was undertaken with the assumption that the CARE tool items can and should have a life beyond the demonstration. The demonstration has shown that the standardization of assessment items across settings is both possible and desirable for a variety of reasons, including more comparable measurement of function and other outcomes, more comparable risk adjustment, and better payment modeling. The demonstration also showed that the collection of patient-specific information in hospital settings such as general hospitals and LTCHs is advisable to better specify differences in the medical, functional, and cognitive complexity of patients treated in these settings.

ES.12 Next Steps

These results have shown what can be done with standardized assessment data. The CARE data are being used in ongoing CMS efforts to further examine some of the similarities and differences among the Medicare population needing physical rehabilitation medicine and those at the other end of the spectrum who may be chronically, critically ill. This work has provided a start to understanding whether similar populations are treated in more than one PAC setting. The results clearly indicate that overlap and substitution exist, although they also highlight that differences in complexity among settings may also be found. Overall, the results highlight the varying characteristics of the Medicare PAC populations and the importance of being able to control for medical, functional, and cognitive status in considering payment reform. More work is needed to develop payment models that will minimize the uncertainty in changing payment systems but improve the consistency of the incentives associated with use across an episode of care.

The CARE items are also being used to consider quality measures. Having standardized measures of case-mix complexity will allow the Medicare program to develop setting-neutral measures that will consistently measure patient outcomes, regardless of site of care.

Standardized items are already being incorporated into the LTCH quality reporting program and are being considered for other measures as well.

Translating the findings presented in this project into actual payment models will require additional work. For example, in future payment projects, two cost components will need further consideration to refine the Medicare payment models. First, further analysis of the patient-specific cost of nontherapy ancillary use is needed to understand how these costs vary by patient complexity. These considerations will be important for determining whether the ancillary costs should be an independent cost component or are highly correlated with any of the medical or functional factors. Current payment approaches for these services that vary by setting will also need to be considered.

Another outstanding cost component is the fixed cost analysis. This demonstration focused on the variable costs associated with patient characteristics. Before designing a unified payment model, the different fixed costs associated with each level of care (e.g., a hospital compared to a nursing facility compared to an HHA) will need to be taken into account. These standard costs can be tied to organizational features, such as size, volume, capital, and other factors that do not vary by patient characteristics and should be considered separate from the variable patient costs.

Additionally, the desirability and feasibility of a composite cost measure that combines the routine, therapy, nontherapy ancillaries, and fixed costs needs to be considered. This report presented analyses of the first two payment components: routine/nursing services and therapy services. Additional payment components, for ancillary service use and for "fixed" setting-specific indirect operating costs, would need to be incorporated to create a complete PPS for the PAC settings. And, ultimately, additional analyses that attempt to link selected outcomes to payment and other incentive structures also will be important.

The results of the analyses in this report demonstrate the importance of including consistent measures of patient medical, functional, and cognitive status in the payment model and of understanding resource intensity variations when considering future PAC PPSs that will optimize patient care while making prudent use of Medicare program/trust fund dollars.